Sujet de master recherche « Architectures logicielles distribuées » 2006-2007

Confidentialité des données dans les systèmes pair à pair

Encadrante principale : Patricia.Serrano Alvarado
Email : Patricia.Serrano-Alvarado@univ-nantes.fr
Co-encadrant : Patrick Valduriez
Email : Patrick.Valduriez@inria.fr

Mots-clés : confidentialité, DHT, pair à pair, bases de données.

Objectif du stage

Ce travail se déroule dans le contexte du projet APPA (Atlas Peer-to-Peer Architecture) [1, 3]. Dans un contexte pair à pair (P2P) comme celui d'APPA et dans les systèmes répartis en général, les données peuvent être collectées facilement et utilisées pour des objectifs divers sans aucun contrôle du propriétaire original. La protection des données est souvent faite à base de rôles (RBAC) mais une fois qu'un individu a accès aux données, il est libre d'en faire ce qu'il veut. Il peut les partager, les utiliser à des fins frauduleuses, pour du marketing, etc. Dans ce contexte, le problème de la confidentialité des données personnelles ou confidentielles émerge naturellement.

En effet, il est de plus en plus nécessaire de protéger les données de manière personnalisée (chaque individu peut avoir des préférences de confidentialité différentes) tout en donnant plus de sémantique aux droits d'accès aux données. Par exemple, un médecin veut bien partager les dossiers de ses patients souffrants d'une maladie rare avec certains de ses collègues mais uniquement pour réaliser une étude bien particulière. Les dossiers seront partagés partiellement afin de préserver les préférences de confidentialité de chaque patient. Ainsi, il y aura certains patients qui ne veulent pas dévoiler leur identité et d'autres pour lesquels cela ne pose pas de problème. Une fois l'étude terminée, les dossiers ne doivent plus être gardées par les médecins collègues.

Dans les systèmes P2P structurés, une DHT (e.g., Chord [6]) est de plus en plus utilisée pour stocker de la méta information, par exemple, un catalogue de données. Ce catalogue peut servir pour le routage de requêtes afin de faciliter la localisation des sources [5].

Dans ce travail nous proposons d'utiliser des requêtes spécifiant les objectifs pour lesquels les réponses seront utilisées [4]. Les sources, de leur part, doivent spécifier les objectifs pour lesquels les données sont partagées. Cette information sera rajoutée sur la DHT afin d'adresser les requêtes, selon leurs objectifs, uniquement aux pairs permettant l'accès aux sources pour les mêmes objectifs.

Travail à réaliser

1. Proposition d'un service de catalogue réparti dans un contexte P2P prenant compte des préférences de confidentialité des sources
2. Développement d'un prototype en Java.
3. Etude des performances par expérimentation et simulation (e.g., SimJava, Brite).

Références

[1] ATLAS-GDD Team at LINA. http://lina.atlanstic.net/fr/equipes/team6/index.html

[2] Rakesh Agrawal, Paul Bird, Tyrone Grandison, Jerry Kieman, Scott Logan, Walid Rjaibi. Extending Relational Database Systems to Automatically Enforce Privacy Policies. Int Conf. on Data Engineering (ICDE), Tokyo, Japan, April 2005.

[3] Reza Akbarinia, Vidal Martins, Esther Pacitti, Patrick Valduriez. Design and Implementation of Atlas P2P Architecture. Global Data Management (Eds. R. Baldoni, G. Cortese, F. Davide), IOS Press, 2006.

[4] Ji-Won Byun, Ninghui Li. Purpose Based Access Control for Privacy Protection in Relational Database Systems. Int. Journal on Very Large Data Bases (VLDB), to appear.

[5] Leonidas Galanis, Yuan Wang, Shawn R. Jeffery, David J. DeWitt. Locating Data Sources in Large Distributed Systems. Int Conf. on Very Large Data Bases (VLDB), Berlin, Germany, 2003.

[6] Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek, Hari Balakrishnan. Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications. In Special Interest Group on Data Communication (SIGCOMM), San Diego, CA, USA, 2001.